

Sponsored since 2011 by the  
*Italian Society for Analytic Philosophy*ISSN 2037-4445 ©  
<http://www.rifanalitica.it>

# Dennett against Cartesian Materialism: the key role of Colour Phi Phenomenon and Multiple Drafts Theory

*Davide Scotti*

**Abstract.** In this article a major argument by D. Dennett is analysed in order to provide an ultimate argumentation against any possible declination of *Cartesian Materialism*, meant as an effective theoretical paradigm to frame any issue concerning the problem of consciousness. In order to do so, Dennett formulates an alternative explanation of consciousness, which is supposedly more consistent with the available empirical evidences concerning the way the brain works. This is the *Multiple Draft Theory*, which the philosopher supports by adopting the *colour phi phenomenon*, a widely known perceptual illusion, as a major supporting evidence. This phenomenon is extremely relevant to Dennett as it challenges a common intuition about conscious experiences, notably instantiated within the *Cartesian Theatre Paradigm* both in its materialistic and dualistic version, namely that the temporal order of conscious experiences reflects the temporal order of the events in the external world that triggered them. The colour phi phenomenon, however, shows that this is not the case, making a reformulation of the basic notions orbiting around the problem of consciousness necessary to be rethought.

**Keywords.** Consciousness, Cartesian Materialism, Multiple Draft Theory, Cartesian Theatre Paradigm, Colour Phi Phenomenon..

## Introduction: Dennett's critical target

The analysis of the colour phi phenomenon represents a fundamental step in Dennett's explanation of Consciousness, as it leads him to definitively give up any legitimacy to the materialistic version of the Cartesian Theatre explanatory paradigm – or *Cartesian Materialism* – left aside its dualistic version. Furthermore, he develops an alternative, more effective explanation for the arising of consciousness, the *Multiple Drafts Theory*.

In W. Seager's words, *Cartesian Materialism* includes "any view that asserts there is some property in the brain which 'fixes' one's states of consciousness and their temporal relationships" (Seager 1999, 111). This condition would be fulfilled by a native theory according to which there is a place in the brain where "it all comes together" and conscious experiences are produced. It would be also fulfilled by a perhaps more acceptable view of consciousness according to which conscious states depend on the global activity of the brain and are just "a matter of a representation exceeding some threshold of activation over the whole cortex" (Dennett 1991, 166). In fact, in both cases we are dealing with a brain mechanism, located in a single place (a so-called neural correlate of consciousness) or spread around the whole cortex (a threshold system), that produces conscious experiences, which share its very same temporal features: the order the stimuli get to the neural correlate of consciousness or make the neurons' activation exceeding the proper threshold corresponds with the order of the conscious experiences caused by these very same stimuli. This is the *Cartesian Materialism* core idea against which Dennett addresses his conclusions concerning the colour phi phenomenon<sup>1</sup>. He claims that the features of our conscious experiences are not mirrored by any brain features, considered as representational vehicles, whereas the only restriction on the relation between a conscious experience and a brain state that underlies it concerns the contents: the latter must somehow underlie the contents of the former, while any other phenomenal property, such as the temporal order of the perceived events in the case of colour phi phenomenon, must not be exhibited by the representational brain states (Akins 1996, 15).

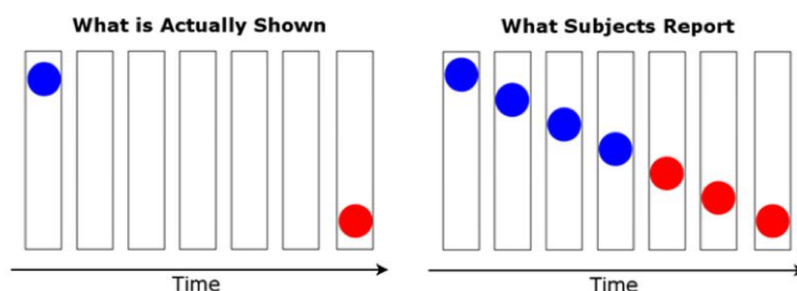
In this article, I will explain how Dennett manages to dismiss any possible version of *Cartesian Materialism* by analysing the temporal properties of the colour phi phenomenon and so providing an alternative explanation of these

<sup>1</sup> *Cartesian Materialism*, as Dennett states, represents a very dangerous temptation for any account of consciousness, independently of their sophistication. For instance, even the eminent Global Neuronal Workspace Model can still be interpreted through the *Cartesian Theatre Paradigm*, although it denies the existence of a neural correlate of consciousness. In fact, if a threshold system or a top-down mechanism was recognised as determining the dominant neural coalition to gain global availability so that consciousness is generated, then there would be a specific brain feature which fixes the conscious experiences temporal order in a Cartesian Theatre-like way (Dennett 2000, 223).

features based on the Multiple Drafts Model.

## 1 Explaining the Colour phi phenomenon: two hypothesis from *Cartesian Materialism*

The colour phi phenomenon was detected first by Michael von Grünau and Paul Kolars in 1976 as a variation of the traditional phi phenomenon. It represents major evidence for the search for object continuity as being a basic feature of visual perception (Seager 1999, 120). Two dots of light are displayed in rapid succession on a screen separated by four degrees of visual angle. The first dot is red and flashes on for 150 msec. Then it goes off and 50 milliseconds later it is followed by a blue dot that flashes on for 150 milliseconds as well. Watching these dots going on and off in such a fashion, a subject would not report to see two different dots flashing in sequence, but he would claim to see one single dot moving and changing colour mid-trajectory instead (Akin 1996, 4).



This phenomenon is extremely relevant to Dennett as it challenges a common intuition about conscious experiences. The intuition is that the order of conscious experiences reflects the order of the events in the external world that triggered them. The colour phi phenomenon, however, shows that this might not be always the case: in fact, it appears that a moving and colour-changing percept of the dot has been first created by the brain and then interpolated between the red and the blue dot. Therefore, as precognition is excluded, the intervening-moving dot seems to be produced retrospectively after the second dot is perceived and then projected backwards in time within the subject conscious experience through a puzzling and mysterious mechanism (Dennett 1991, 6). At the very least, this phenomenon reveals that, at a micro-time scale, the temporal order of the events included in our conscious experiences is determined by unconscious cognitive processes and it does not need to reflect the actual order of the correlated events in the external world. For this reason, the phi phenomenon appears to Dennett as a possibly powerful evidence against *Cartesian Materialism*, since the latter states that there should be a fixed criterion to determine the temporal order of

conscious experiences (Seager 1999, 120). Starting from these considerations, Dennett engages in a confrontation between two possible explanations of the phi phenomenon which are ascribable to *Cartesian Materialism*, on the one hand, and his own explanation, which is based on the Multiple Drafts Model, on the other hand. Dennett's aim here is to disengage the features of the represented events - those in the external world coded within the brain processes underlying consciousness - from the features of the phenomena representing them, namely the subject's conscious experiences (Dennett 1992, 9). His strategy aims at denying any mechanism which fixes the latter according to the former, so that *Cartesian Materialism* would be definitively falsified.

Dennett considers two possible explanatory hypothesis of the colour phi phenomenon which could be put forward as defenders of the *Cartesian Theatre Paradigm*: the Stalinesque Revisionist Hypothesis (H1) and the Orwellian Revisionist Hypothesis (H2).

According to H1 the conscious experience of the moving dot is generated after the second dot is presented. Before that, the perceptual inputs of the two dots are revised so that inferences about those events are made and other illusory perceptual events are created and interpolated between the two dots in the manufacturing of the final perceptual experience.

According to H2, however, the conscious experiences of the two distinct dots occur first, then the same steps postulated by H1 take place: the same inferences result in the same conclusion and an illusory visual state is manufactured. Then, the memories of the first experiences are erased and substituted by the new illusory memories which consequently enter consciousness domain (Akin 1996, 5-7).

The main difference between these two models concerns the moment within the sequence of brain processes when the perceptual content becomes conscious: in the Stalinesque model this happens only in the end, when the perceptual experience is manufactured, while in the Orwellian model it happens twice, immediately after the input is received and after the false memories about the perceptual experience substitute the original memories (Akin 1996, 8). As Dennett states, however, this distinction is actually not relevant, as H1 and H2 are impossible to be distinguished experimentally: on the one hand, such a precise serial sequence of processes cannot be discriminated observing the brain activity, while, on the other hand, by probing the subject, it is impossible to differentiate a revised perceptual experience from a revised memory of it at such a temporal scale. According to Dennett, in fact, probing subjects experiencing the colour phi phenomenon just proves that what they refer to, when they talk about their conscious experiences actually consists of fake perceptual memories and nothing more. However, it cannot be known if these memories are memories of an illusory conscious experience (H1), or if they substitute the original memories of a not

illusory conscious experience (H2) (Dennett 1992, 17).

Therefore, Dennett's argumentation can be summarized in this way: confronting H1 and H2, he uses a verificationist argument to claim that, as there is no way to distinguish them, there is no difference between them<sup>2</sup>; then he opposes the core Cartesian perspective shared by both – that there are conscious states whose features are somehow preliminarily fixed according to univocal brain mechanisms (1) – with the claim that there is no conscious states (2) but only false memories of conscious experiences continuously formed (3); he eventually concludes, consistently with a verificationist strategy, that the first claim (1) and the first part of the second claim (2) are undistinguishable as unprovable, while only the second part of the second claim (3) is verified by the subjects' probes (Seager 1999, 123).

This conclusion deeply threatens the existence of consciousness as conceived by *Cartesian Materialism*. The reason is that, what turns out as conscious, namely the reported fake memories, does not fulfil the necessary condition required by *Cartesian Materialism*: the temporal order of these conscious experiences does not appear to be fixed according to a univocal mechanism within the brain. However, this is not enough to eliminate any possible version of *Cartesian Materialism* from the set of the possible explanations of the colour phi phenomenon. In fact, consciousness can still be conceived as a product of brain states which is generated within it at some specific moment, even if it is made of fake memories and it doesn't mirror the temporal features of the outside-world events. Therefore, a further move is needed in Dennett's argumentation in order to definitively tackle the core idea of the *Cartesian Theatre Paradigm*. This move will be made by providing the already verified fact that only false memories are somehow conscious with an adequate explanation, which is empirically grounded and, at the same time, not consistent with the *Cartesian Theatre Paradigm*. The Multiple Drafts Model will serve the cause.

## 2 Dennett's target: the Cartesian core of Stalinesque and Orwellian Revisionist Hypothesis

In the first paragraph I have analysed the explanation of the colour phi phenomenon provided by H1 and H2 and Dennett's refusal of their legitimacy based on a verificationist argument. I will now consider more deeply the core idea underlying both H1 and H2, in contrast to which Dennett's alternative explanation of consciousness based on the *Multiple Drafts Theory* arises.

<sup>2</sup>Seager names Dennett's verificationism "micro-verificationism" since he restricts its application to short periods of time, namely a few hundred milliseconds. This, in fact, is the time scale employed in measuring neurophysiological events. The core principle, anyway, remains the same: if something cannot be verified, it is not true (Seager 1999, 122).

The Stalinesque and the Orwellian Hypothesis are two alternative explanatory models equally based on the same conception of the brain processes underlying conscious experiences – or, more precisely, what the subject would report as such – as a serial sequence of operations within the brain. According to this fundamental idea, the colour phi-phenomenon occurs since, after the perception of the two inputs – which could be either conscious (H1) or unconscious (H2) – a series of propositional inferences led by the urge for object continuity takes place, and results in the conclusion that the dot has moved and has changed its colour. As a consequence, a conscious perceptual experience (H1) or a false memory of it (H2) is manufactured accordingly (Akin 1996, 7). The seriality of these processes is the crucial feature that allows the identification of a turning point, after which the contents of the brain processes become conscious. Therefore, this very moment in the process can be conceivably identified in order to provide a criterion for the order of conscious experiences to be established consistently with the order of the external events provoking the stimuli. This is exactly the core idea underlying the Cartesian Theatre metaphor criticised by Dennett, and it occurs in both the Stalinesque and the Orwellian models in the form of a further representation – a revision indeed – of the stimuli that have already been discriminated once and are thus already coded in contentful brain processes (Dennett 1992, 4). In fact, according to these models, the input from the external world, after it has been discriminated through several lower level brain processes, is supposedly represented as a contentful brain state to another brain system in order to become part of the subject's conscious experience. In the Stalinesque model this happens when the brain produces a perceptual conscious experience after having discriminated and revised all the visual inputs provided by the display of the colour phi phenomenon. In the Orwellian model this happens even twice: in the first place when the first perceptual experience is produced and in the second place when the revised memories substitute the storage memories of the first visual experiences (Akin 1996, 8).

### **3 The *Multiple Drafts Theory*: an alternative explanatory model**

In order to dismiss the already unverifiable explanatory models based on the *Cartesian Theatre Paradigm*, Dennett describes the way brain mechanisms might function in parallel, providing so a convincing explanation of any possible colour phi-kind of phenomena and consequently excising overall the idea of serial processes within the brain. This parallel-processing system is named Multiple Drafts Model and it relies on well-grounded experimental studies. Its core idea is that “all varieties of perception – indeed all varieties of thought or mental



activity – are accomplished by parallel, multi-track processes of interpretation and elaboration of sensory inputs” (Dennett 1991, 111). Therefore the brain is a parallel processing machine, so that any function implemented in it, also the so-called consciousness, relies on mechanisms based on parallel processing.

What is mainly relevant in this model for Dennett’s theoretical purposes is the denial of a conceivable precise moment or step within the overall process which can be considered as the when and where of conscious states generation, a sort of conscious experiences output. On the contrary, there is not supposed to be any fixed pathway or mechanism that determines univocally which brain contents become conscious and which do not and when they do. Conversely, the brain processes configuration underlying each conscious state reflects the overall activity of the brain and it is always different according to the specific conditions of each moment, such as the hardware constraints, the kind of stimuli involved in the processes, the environmental conditions or the subject’s behavioural demands.

According to the Multiple Drafts Model, the brain is constituted of several modules. Each module has a specific function and it is interconnected with many other modules in a not-hierarchical way. Most of the connections are reciprocal and different processes run simultaneously. Within this model a visual stimulus, for instance, would initially pass through the lowest levels of the visual system, where the basic features of the receptive field related to each retina cell, which are not part of our conscious experience, are serially detected (LGN, V1 and V2). After that, it would spread among several modules which process different higher-level information about it simultaneously. Each of these modules would process and exchange with each other different kind of information about the very same stimulus: shape, colour, position, movement are some of its properties discriminated by these modules. By communicating with each other, they manage to jointly generate several more or less effective alternative interpretations of the stimuli, competing with each other to provide the whole system with a representation of the external world: these are the so-called “multiple drafts” (Akin 1996, 13). One of these interpretations, resulting from the work of a certain configuration of parallel brain processes, overcome the others, that involve different configurations, and it achieves the monopoly over the brain resources in that very moment. Consequently, it determines the whole system behavioural outcome, which includes also what the subject perceives, constituting thus a moment of the subject’s conscious experience. No top level control is needed, but a kind of “democratic” competition among different configurations takes place (Dennett 2001, 225).

In this overall process each module determines certain properties of the stimuli by ordinary computational process, whose conclusion can be theoretically timed (Akin 1996, 13). The conscious states generation process, however, cannot be timed, since such a process is not needed at all. In fact, there is no manu-

factured consciousness, but only several processes with different functions and different timings, covering different brain pathways: they jointly determine the commonly called conscious states, which are nothing else but the combined effect of these interacting brain processes on the whole system. In the colour phi phenomenon case, for example, the two stimuli are processed by the brain through several spread parallel processes which represent different features of them. From the interaction among these processes, a possible interpretation about the relation between the two dots emerges, whose effect is the subject experiencing the one moving and colour-changing dot (Akin 1996, 20-23).

Therefore, since the criteria for the dominant interpretation to win this competition among several “drafts” are not fixed nor predictable and thence there is no proper consciousness generation process through which the contents of conscious experiences are sequentially determined, the temporal features of a conscious experience do not have to reflect the temporal features of the underlying contents-carrying brain processes<sup>3</sup>: there is no trace of *Cartesian Materialism* eventually.

#### 4 Consciousness re-conceived: Cerebral Celebrity and Clout

So far I have shown how the Multiple Drafts Model is able to offer an alternative explanation of the colour phi phenomenon, which, as Dennett states, deprives any declination of *Cartesian Materialism* of explanatory efficacy. Therefore, since *Cartesian Materialism* conceptualisation has to be set apart together with its account of consciousness, Dennett provides a new way to conceive the concept of “consciousness” by using two metaphors to describe it: thus, consciousness is described as a “cerebral celebrity” within the brain, or it can be explained in terms of “clout”.

As stated in the previous paragraph, according to the Multiple Drafts Model, what makes a configuration of interacting, contentful brain processes conscious is its temporary control of the whole system. On the strength of this consideration Dennett compares consciousness to a brain process - or a configuration of brain processes - which is a trending celebrity among all the brain processes at a certain moment: this brain process would then hold the power to control the subject’s behaviour in virtue of its fame around the whole brain processes community (Seager 1999, 113). This metaphorical representation of consciousness is intended to contrast the intuitive and shared idea of consciousness as a “medium of representation” within the brain “into which content-bearing events must be

<sup>3</sup>The Multiple Drafts Model of the brain seems to boast some relevant empirical supporting evidence, such as D. Van Essen’s macaque visual system model (Akin 1996, 9).



transduced in order to become conscious”, a view that is strictly committed to the *Cartesian Theatre Paradigm* (Dennett 2000, 224). Conversely, consciousness is conceived as being what happens when an unconscious contentful brain process among several others competing with each other achieves enough “clout” to temporarily dominate the “political scene” within the brain and consequently control the body.

Consciousness, understood as anything that is included in the subject’s experience, is then a sequence of dominant coalitions of brain processes which follow one another according to the very moment conditions and demands of the environment (Dennett 2000, 225).

The distinct concepts of “fame” and “clout” are used seemingly here, but, even if they are strictly connected, they actually underline two different aspects of conscious states within the metaphorical framework of the Multiple Drafts Model. The concept of “Fame”, in fact, endorses the concept of knowledge and lay emphasis on the content of this knowledge, while “clout” refers to the power to control others’ behaviour. Somebody is famous if they are known by many people, and the more they are known the more influential they are on these many people. The way other people are influenced by celebrity depends on the way these people consider the holder of this celebrity. Similarly, within this brain model, “celebrity” or “fame” refers to the amount of consideration into which the content of a certain brain processes configuration, namely the combination of certain properties of the stimuli processed by each neural mechanisms involved, are taken by the rest of the system. Then, the “clout” of the dominant coalition of brain processes follows from its fame around the brain, so that the fame of the coalition is proportional to its power to determine the whole system behavioural outcome. However, the way this clout is exercised in order to influence the body behaviour is determined by its content.

## 5 Temporal properties of conscious experiences

This distinction between “celebrity” and “clout” sheds light on Dennett’s characterization of conscious states as narratives (Seager 1999, 120). In fact, he also defines consciousness as a fiction, a narrative that is continuously edited, revised and modified by the brain, whose content, moment by moment, consists of judgements concerning the features of the external world, provided by the brain modules included in the same dominant coalition dealing with the interpretation of the incoming stimuli. Each module contributes to the formation of an overall interpretation of the stimuli, and the more an interpretation gain support from other brain modules - that is the more famous it is- the more clout on them it has. The resulting interpretation, then , is what determines the way the clout is exercised, namely the narrative content of conscious experience.

Such a way to conceptualize consciousness is useful to disengage it from the dangerously obvious idea that the order of the experienced events has to be mirrored by the order of the corresponding events in the external world. As it should now be clear, the reason is that timing has been recognised as a matter of interpretation of the stimuli provided by the dominant brain processes coalition. In fact, according to the Multiple Draft Theory, the temporal properties of conscious experiences are conveyed as time-related judgements within the interpretation of the stimuli received by the brain, independently from the actual order these stimuli are received by the system. To clearly understand this, it might be useful to draw an analogy with language. Indeed, in a proposition describing a certain series of events, the order of its parts does not need to mirror the order of the described events, but the timing of the represented events can be rather conveyed otherwise, such as by adverbs. In the same way, information about timing are possibly conveyed symbolically by the brain states underlying conscious experiences, so that they do not need to reflect the order of the external events which triggered them (Dennett 1992, 35). This is normally not true at a macroscopic time scale - the time scale we would say that fits our experience – as in this case there is usually a correspondence between the events in the external world and the events in our conscious experience and it is accepted by anybody that the order of the first ones fixes the order of the second ones. At a micro-time scale, however, this correspondence is not necessary, since an interpretation of certain stimuli whose features diverges from the those of the actual events provoking them can get enough clout over the rest of the system to determine the subject's perceptual experience and thus drive his behaviour. According to Dennett, this is exactly what happens in the colour phi phenomenon: the network of brain processes carrying the interpretation of the stimuli as provoked by a moving- and colour changing-dot takes control over the whole system, determining the experience reported by every subject. This might happen because evolution privileged somehow the ability to detect continuity in moving object over the accuracy of temporal representation of stimuli. Therefore, at the colour phi phenomenon time scale, the timing of the experienced events is supposed to be coded in a symbolical way by the brain processes which are about these events, exactly as adverbs in natural language propositions or the practice of dating letters can be used to symbolically locate the events they refer to in time (Dennett 1992, 9).

## **6 Conclusion: from *Cartesian Materialism* falsification to consciousness re-conception**

I will now schematically summarize Dennett's argument against *Cartesian Materialism*, underlining the crucial role of colour phi phenomenon explanation.

The overall argument can be summarized as follows:

$A \supset B$

$B \supset C$

$C \vee D$

$E \supset D$

$E$

$D$

$\neg C$

$\neg B$

$\neg A$

In order to falsify B Dennett analyses the Stalinesque -H1- and the Orwellian -H2- explanatory models of the phi phenomenon which fulfil the sufficient condition B to be considered as proper declination of Cartesian Materialism ( $B \rightarrow C$ ). He establishes an exclusive disjunction between H1 and H2 core idea (C) and the opposite idea that the conscious experiences temporal order (at a micro-time scale) is not fixed accordingly to the underlying brain states temporal order ( $C \vee D$ ). He uses a verificationist argument to show how H1 and H2 are undistinguishable as unverifiable, while the only empirical fact about the colour phi phenomenon, namely that conscious experiences consist of false memories, is better explained by the Multiple Draft Model (E). This alternative explanation implies D, thus, once Dennett assumes E, he derives D, then  $\neg C$ ,  $\neg B$  and eventually  $\neg A$ : Cartesian Materialism is falsified.

Once Cartesian Materialism is falsified, the way to conceptualise consciousness which has originated from it must be set apart, and the concept of “consciousness” must be rethought. Dennett thus dismisses the idea of consciousness as something that is produced within the brain in a certain moment through a certain system, and endorse instead the idea that consciousness, meant as first person experience, is spread around the brain, parcelled out to various brain modules and “incorporated into the multifarious further effects of all the political influence achievable in the competitions” among several parallel brain processes (Dennett 2000, 228). The “cerebral celebrity” and “clout” metaphors support this re-definition purpose, connecting the definition of consciousness to the observable manifestation of its effect, whose colour phi phenomenon perceptual experience is just an example.

## References

- Akins, Kathleen (1996). "Lost in the Plot? Reconstructing Dennett's Multiple Drafts Theory of Consciousness". In: *Mind & Language* 11.1, pp. 1–43.
- Chalmers, David J. (1995). "Facing Up to the Problem of Consciousness". In: *Journal of Consciousness Studies* 2.3, pp. 200–219.
- Dennett, Daniel (1991). *Consciousness Explained*. New York: Back Bay Books/Little, Brown and Company-Hachette Book Group.
- (2000). "Are we explaining consciousness yet?" In: *Cognition* 79, pp. 221–237.
- Dennett, Daniel and Marcel Kinsbourne (1992). "The Time and the Observer: the Where and When of Consciousness in the Brain". In: *Behavioural Brain Sciences* 15, pp. 183–247.
- Gallagher, Shaun (2000). "Philosophical conceptions of the self: implications for cognitive science". In: *Trends in Cognitive Sciences* 4.1, pp. 14–21.
- Hohwy, J. (2016). "The self-evidencing brain". In: *Noûs* 50.2, pp. 259–285.
- Northoff, G., P. Qin, and T.E. Feinberg (2011). "Brain imaging of the self. Conceptual, anatomical and methodological issues". In: *Consciousness and Cognition* 20.1, pp. 52–63.
- Seager, William (1999). *Theories of Consciousness: An Introduction*. London-New York: Routledge.
- Van Gulick, Robert (2016). *Consciousness*. Ed. by Edward N. Zalta. The Stanford Encyclopedia of Philosophy. URL: <http://plato.stanford.edu/archives/win2016/entries/consciousness/>. Winter 2016 Edition.